

Many-Channel DMA Virtualization for PCIe®-Connected Enterprise SoCs

Stephane Hauradou

CTO

PLDA



Copyright © 2017 PCI-SIG® - All Rights Reserved

Disclaimer

<u>Presentation Disclaimer</u>: All opinions, judgments, recommendations, etc. that are presented herein are the opinions of the presenter of the material and do not necessarily reflect the opinions of the PCI-SIG®.

Copyright © 2017 PCI-SIG® - All Rights Reserved

Agenda

- **Introduction**
 - PCIe® Virtualization and the need for efficient DMA
- **Traditional DMA**
 - Architecture, limitations
- **The vDMA Virtualized DMA Engine**
 - Requirements, architecture, performance, implementation details
- **Q&A**

Copyright © 2017 PCI-SIG® - All Rights Reserved

Driving Factors

- **Exponential traffic growth in data centers**
- **More connected devices, virtualization expanding**
- **Convergence of Network, Compute, Storage on chip**

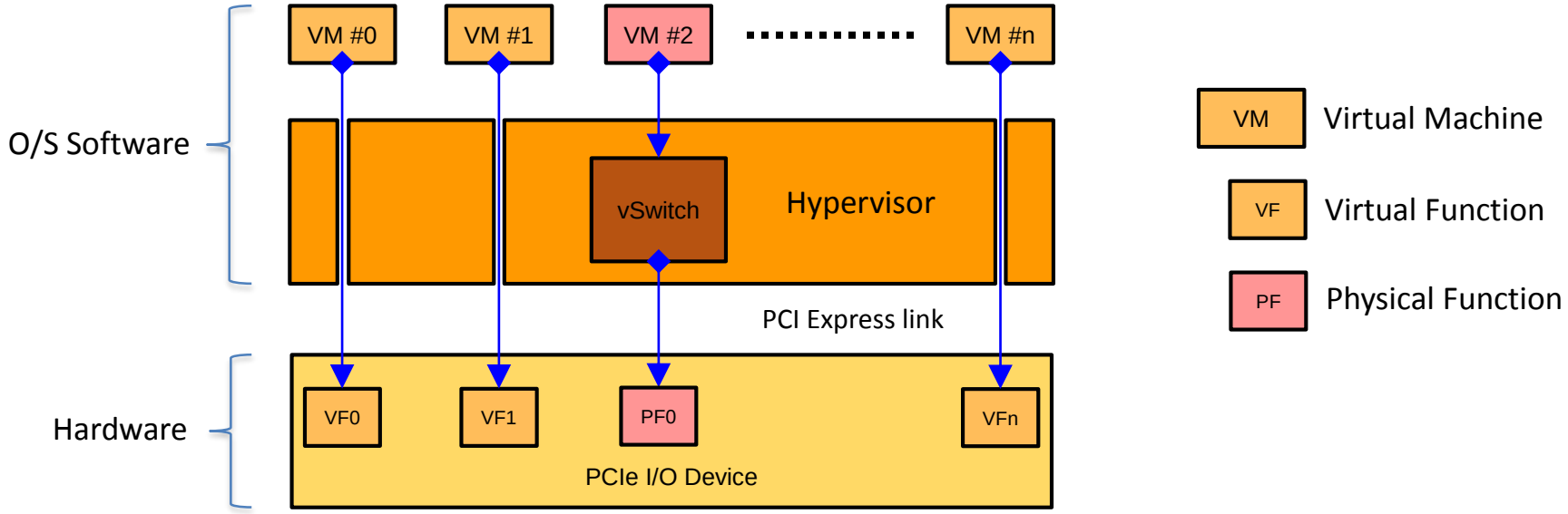


1000s of VMs competing for the same hardware

Copyright © 2017 PCI-SIG® - All Rights Reserved

About PCIe Virtualization

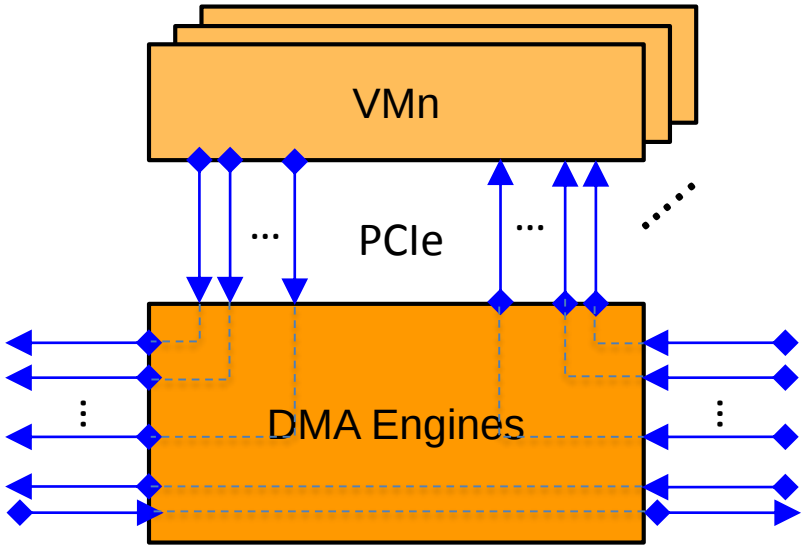
- **PCI-SIG PCIe Specification Extension**
 - SR-IOV, MR-IOV
- **Allows multiple VMs to share an I/O device**
 - Bare metal performance through Hypervisor bypass



Copyright © 2017 PCI-SIG® - All Rights Reserved

VMs Need Efficient Data Movers

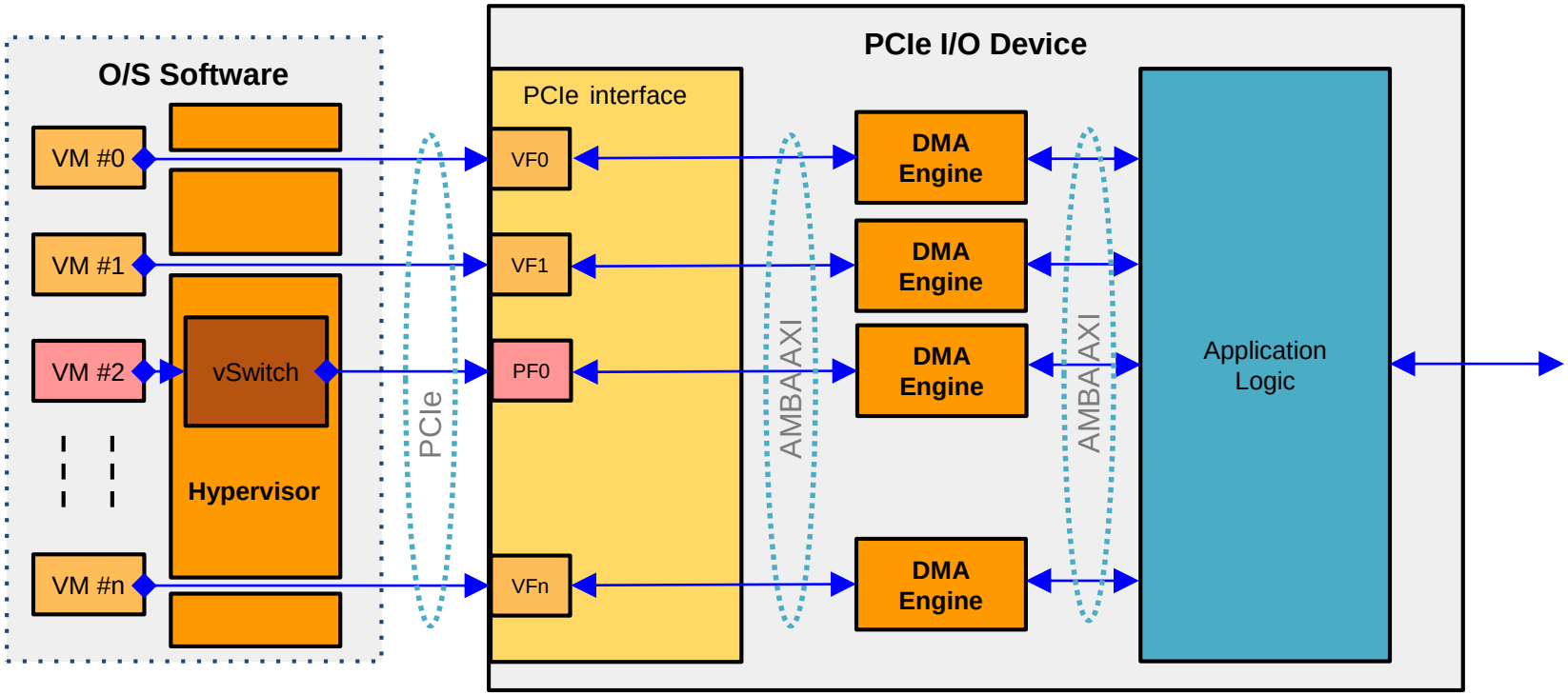
- **<u>Flexibility</u>: Many channels, any direction**
- **<u>Serviceability</u>: Fair bandwidth sharing, non-blocking**
- **<u>Reliability</u>: Isolation, data protection**



Copyright © 2017 PCI-SIG® - All Rights Reserved

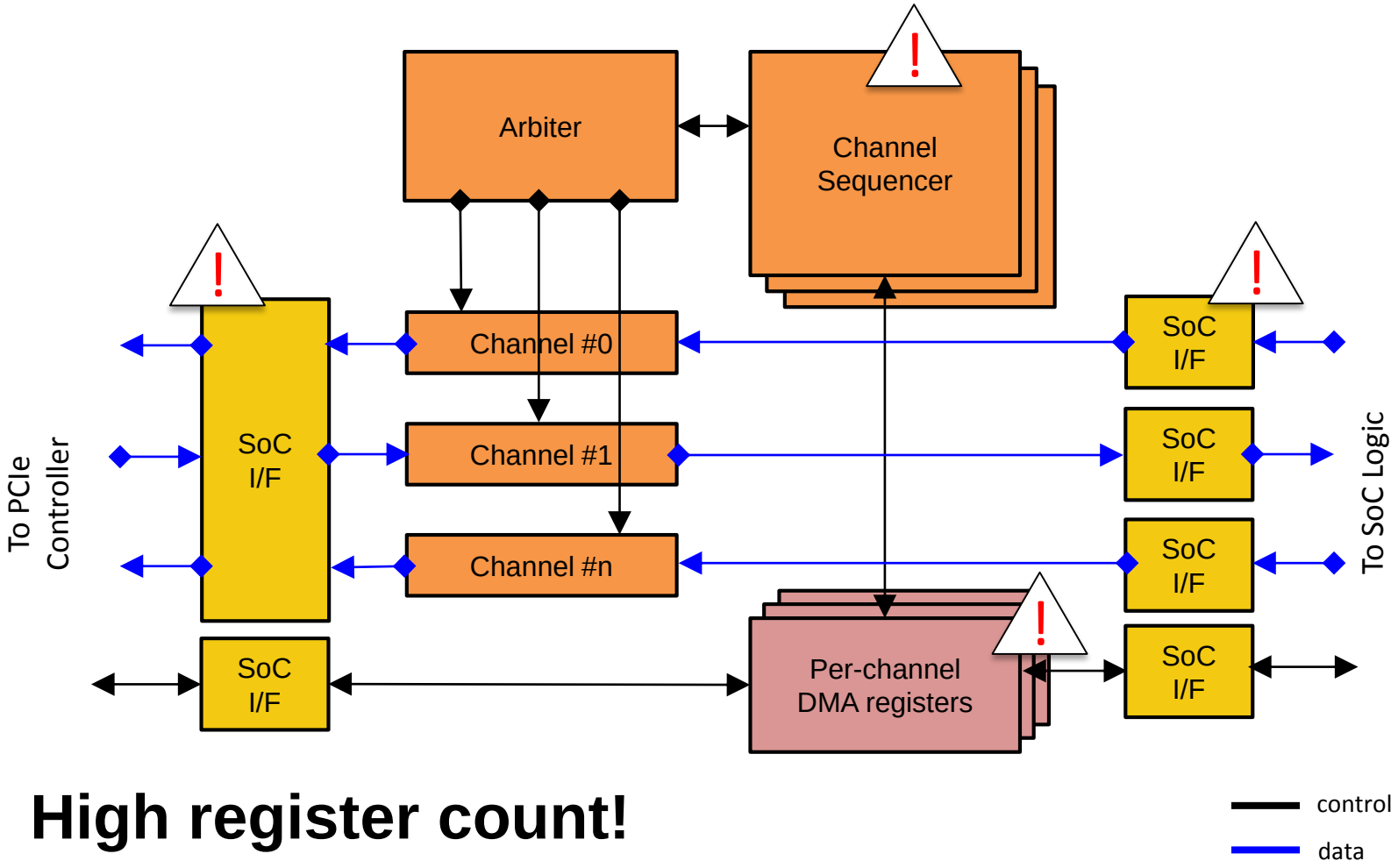
In an Ideal World…

- **Each VM has its own DMA resources**
- **N-channel DMA per VF**



Copyright © 2017 PCI-SIG® - All Rights Reserved

Traditional DMA Architecture

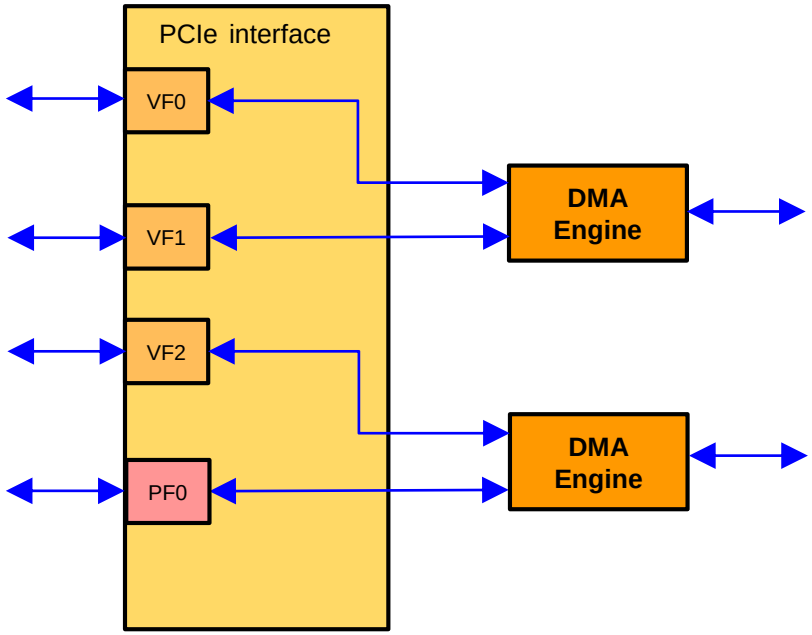


- **High register count!**

Copyright © 2017 PCI-SIG® - All Rights Reserved

IOV DMA Sharing

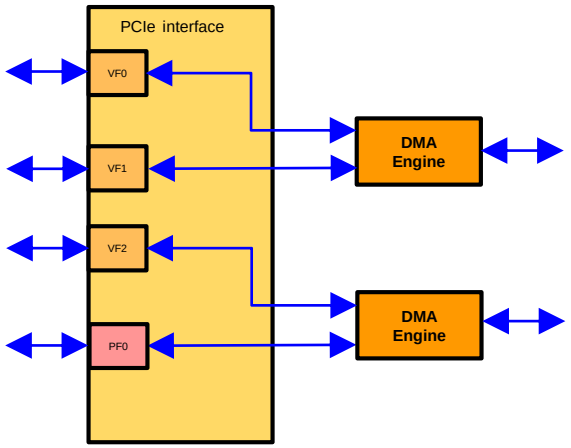
- **Allows for more VF/VMs**



Copyright © 2017 PCI-SIG® - All Rights Reserved

IOV DMA Limitations

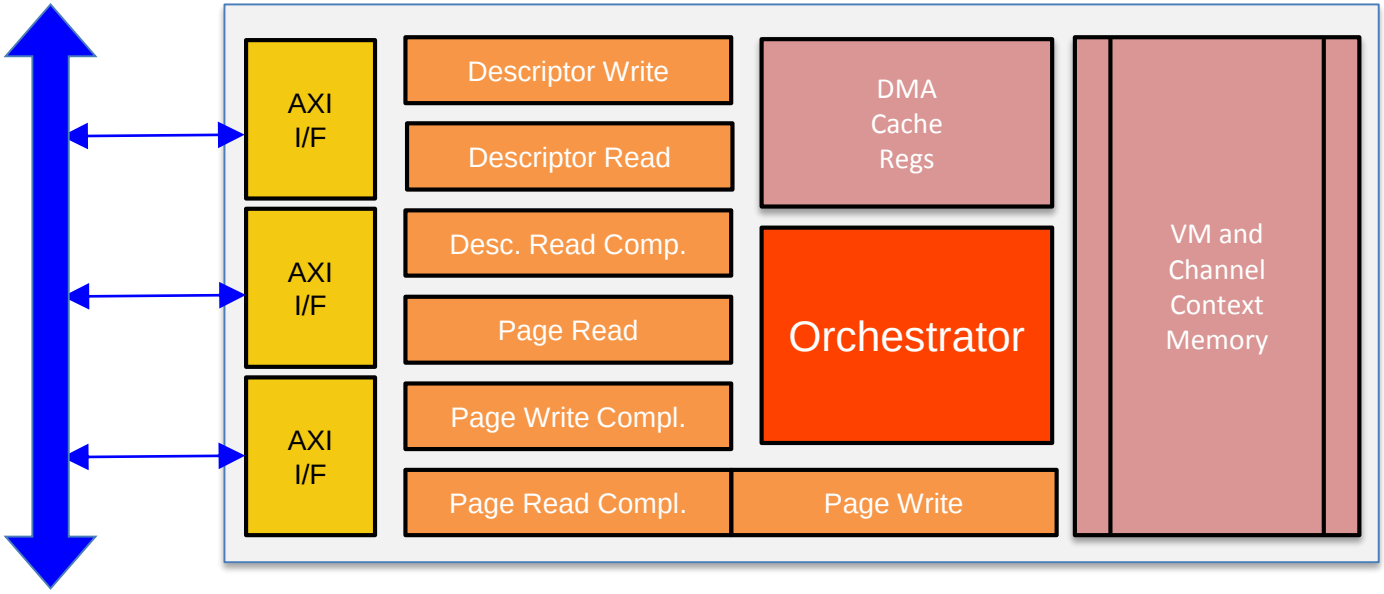
- **VF context switching latency**
- **VF blocking**
- **VM isolation**
- **Scalability**



Copyright © 2017 PCI-SIG® - All Rights Reserved

vDMA IP Architecture

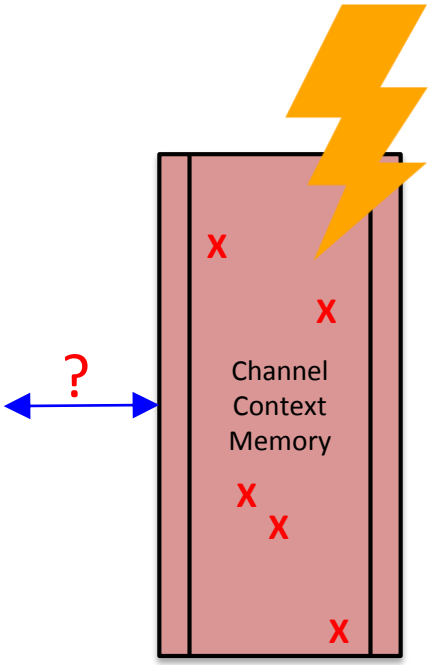
- **A novel approach with:**
 - Concurrent DMA tasks
 - Centralized Context memory
 - Cache registers

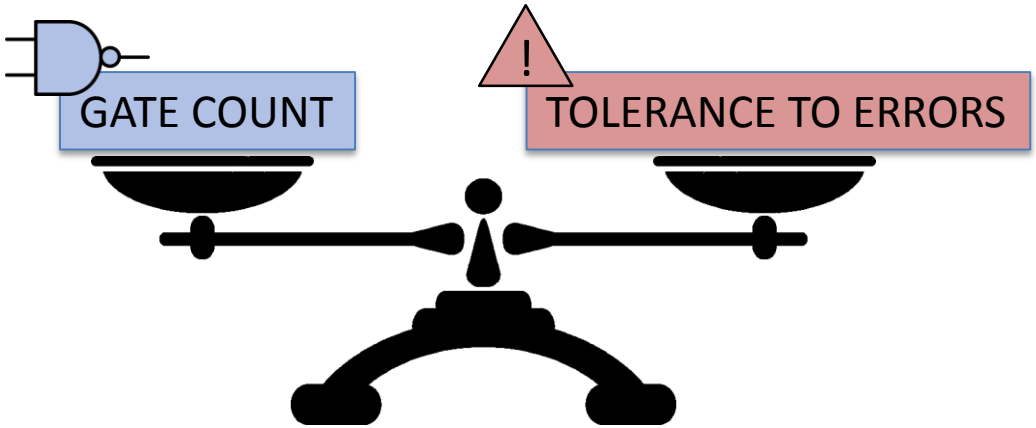


Copyright © 2017 PCI-SIG® - All Rights Reserved

Context Memory Considerations

- **Parity, checksum, CRC**
 - Report errors to VM
- **ECC**
 - Only report NC errors to VM

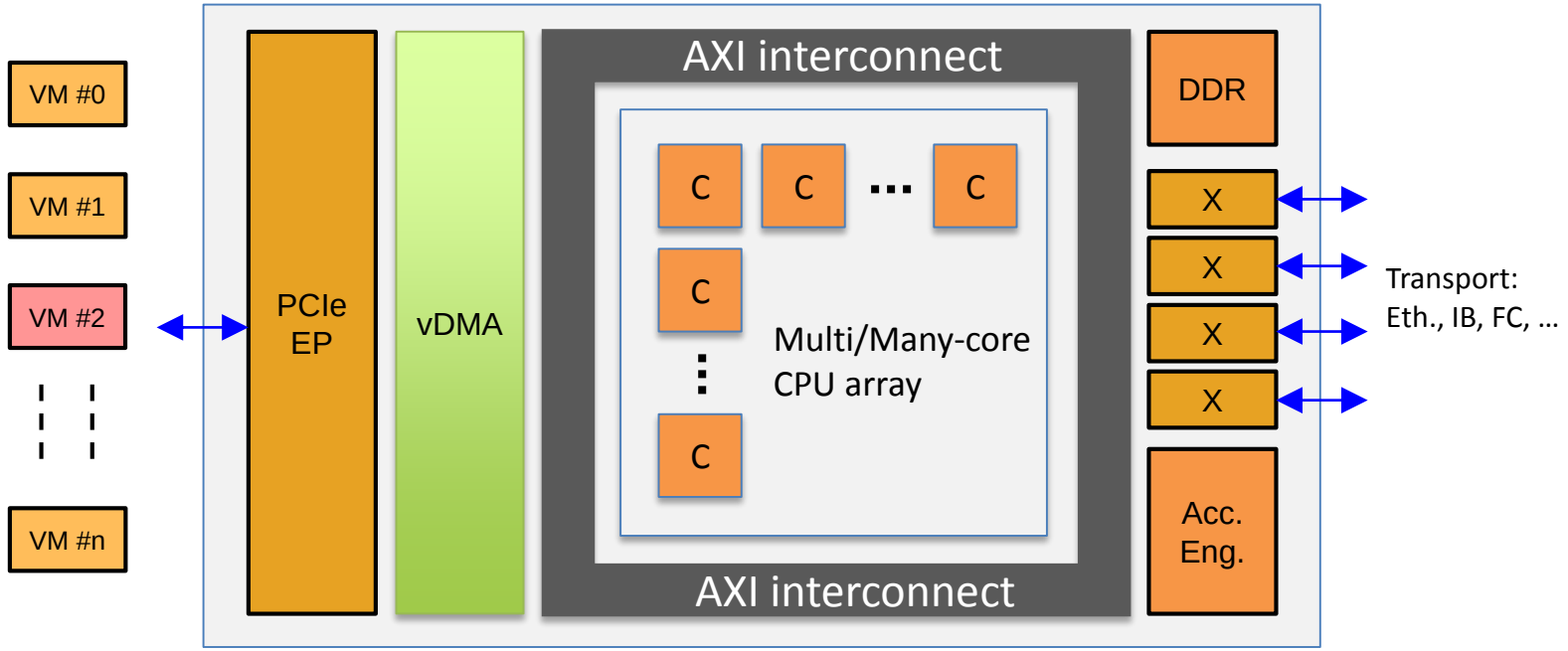




Copyright © 2017 PCI-SIG® - All Rights Reserved

vDMA Use Models (1)

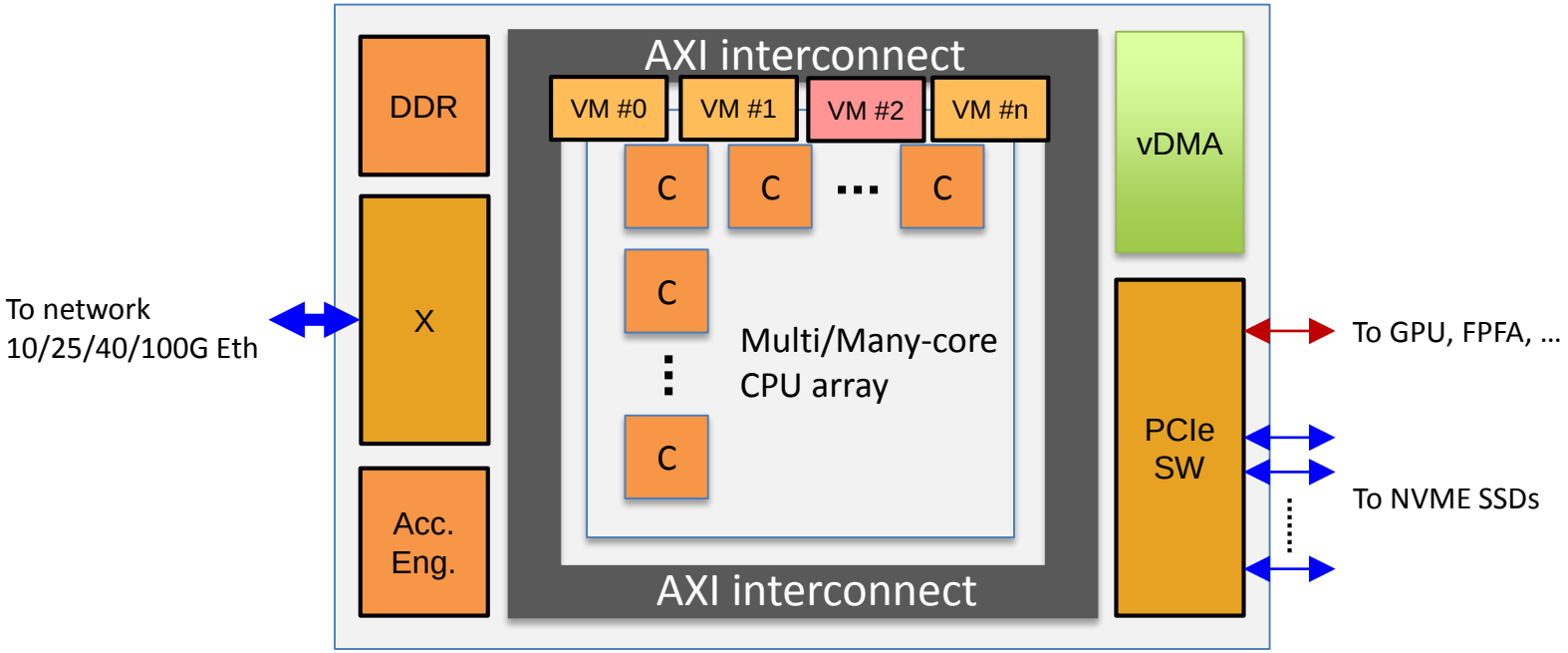
- **Data Center SoCs for Network & Storage offload**
 - Flow processors, SmartNICs, Storage controllers



Copyright © 2017 PCI-SIG® - All Rights Reserved

vDMA Use Models (2)

- **Data Center SoCs for Network & Storage offload**
 - NVMe Storage controllers



Copyright © 2017 PCI-SIG® - All Rights Reserved

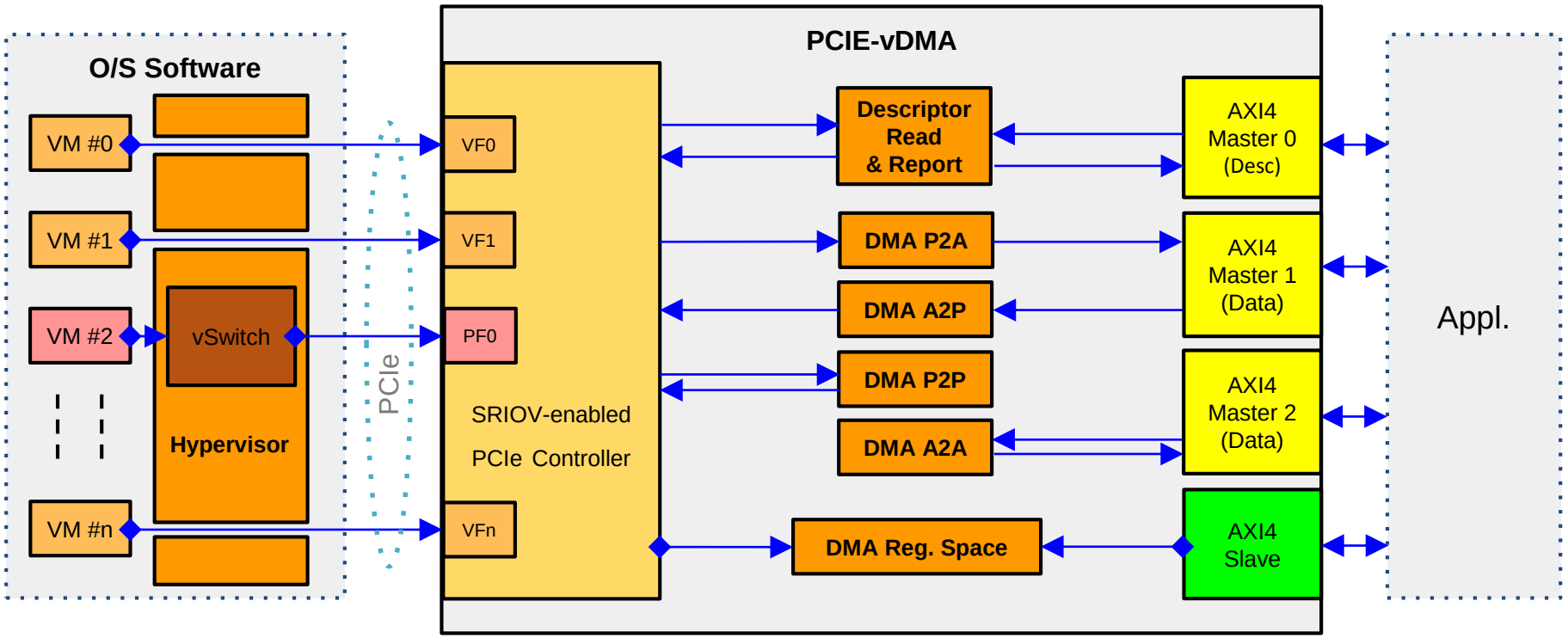
vDMA Requirements

- **Performance & Scalability**
 - High throughput esp. small packets
 - 32+ DMA channels per VM
- **QoS**
 - Non-blocking operations
 - VM and channel priority
- **Security**
 - VM isolation, data path protection



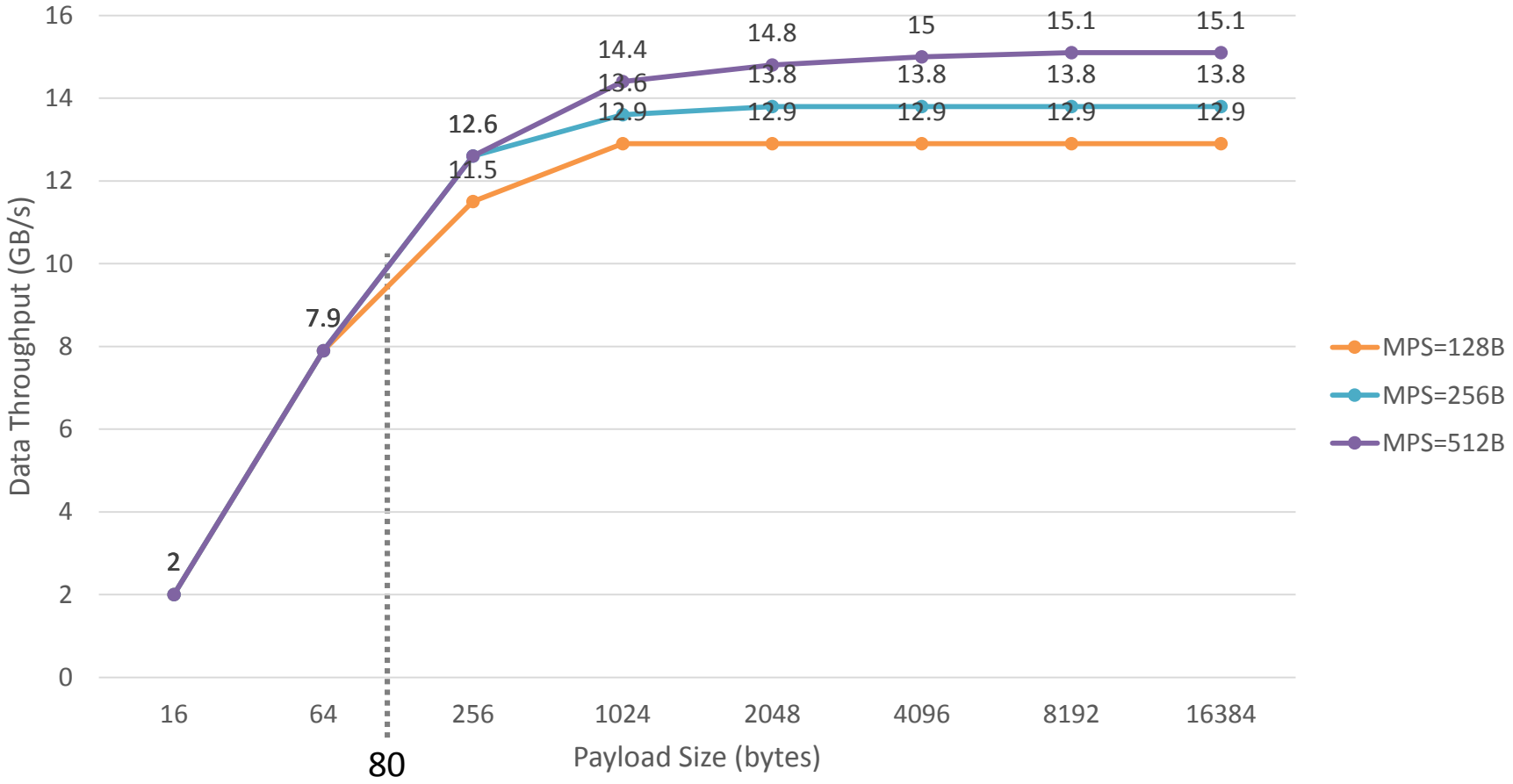
Copyright © 2017 PCI-SIG® - All Rights Reserved

vDMA Data Path



Copyright © 2017 PCI-SIG® - All Rights Reserved

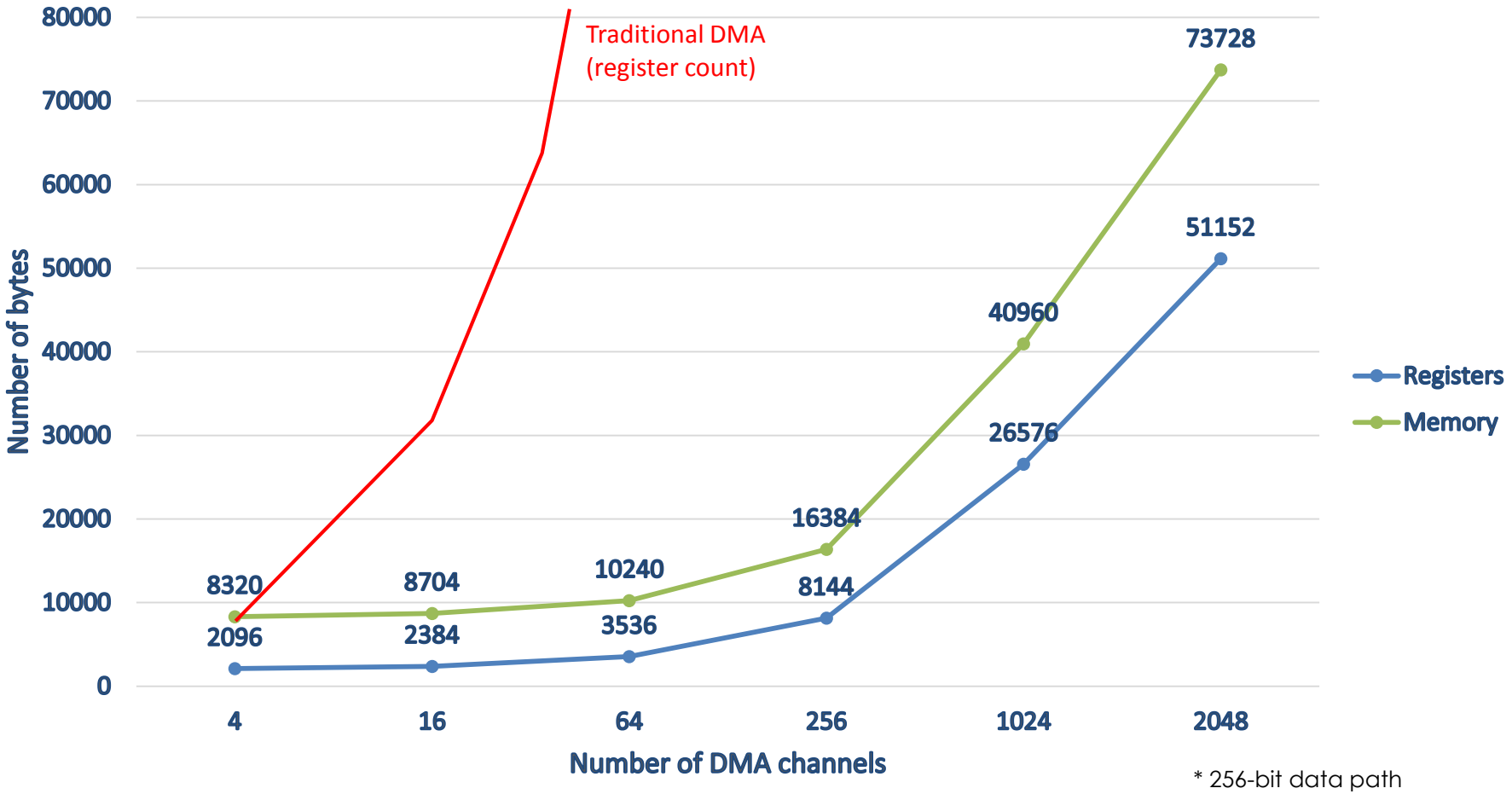
vDMA Performance



* PCIe x16 Gen3
* 128 Outstanding Read Requests

Copyright © 2017 PCI-SIG® - All Rights Reserved

vDMA Resources

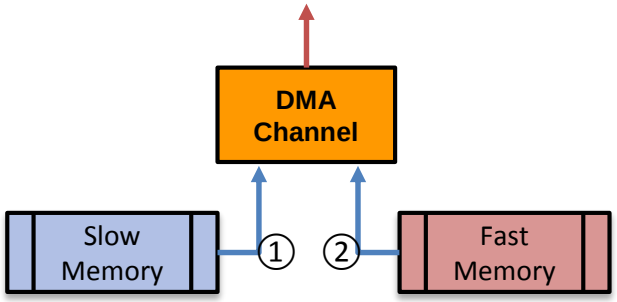


* 256-bit data path
* Excludes PCIe controller logic

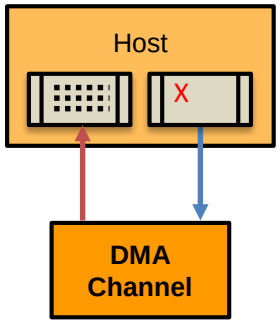
Copyright © 2017 PCI-SIG® - All Rights Reserved

Technical Implementation
Non-Blocking Operation

Channel latency mismatch



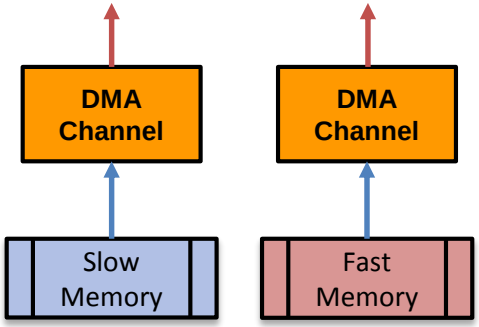
Host buffer/data unavailable



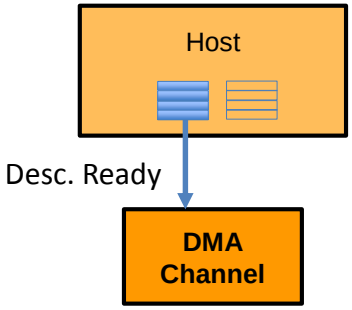
VM crash



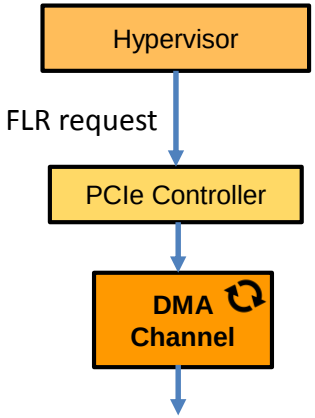
Separate channels



Host-to-DMA protocol



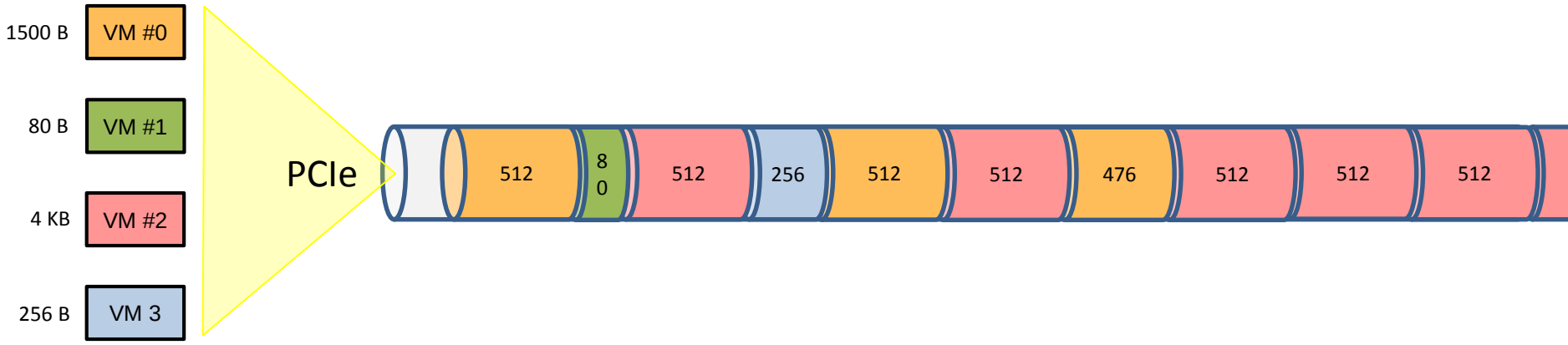
Channel reset via FLR



Copyright © 2017 PCI-SIG® - All Rights Reserved

Technical Implementation
Fair Arbitration Between VMs

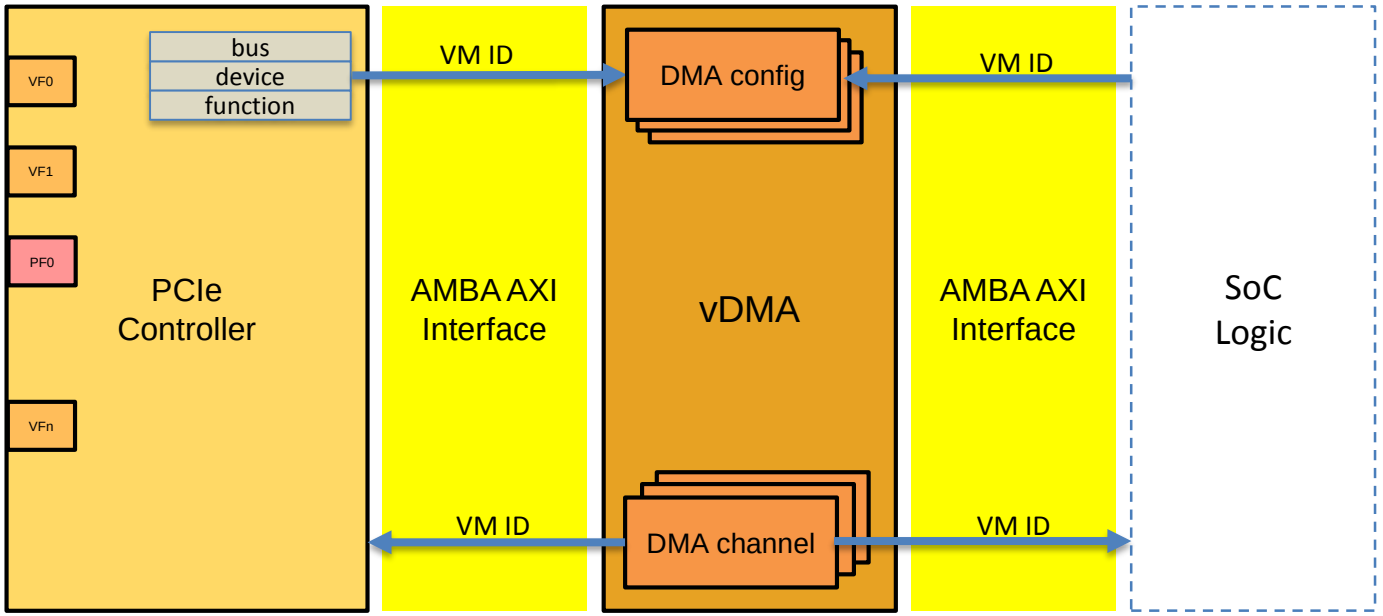
- **Round-robin bandwidth sharing**
 - Fixed packet size, statically configurable



Copyright © 2017 PCI-SIG® - All Rights Reserved

Technical Implementation
VM Isolation

- **Use of sideband signaling**



Copyright © 2017 PCI-SIG® - All Rights Reserved

Technical Implementation
High Throughput for Small Packets

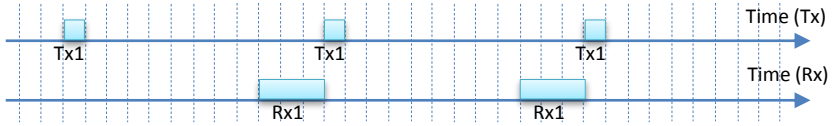


Fig. 1 - 1 outstanding request, small packets

$$\mathbf{BW} = \frac{\boldsymbol{S} * \boldsymbol{N}}{\boldsymbol{RTL}}$$

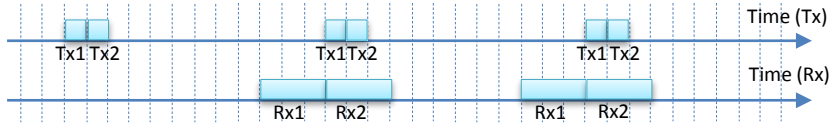
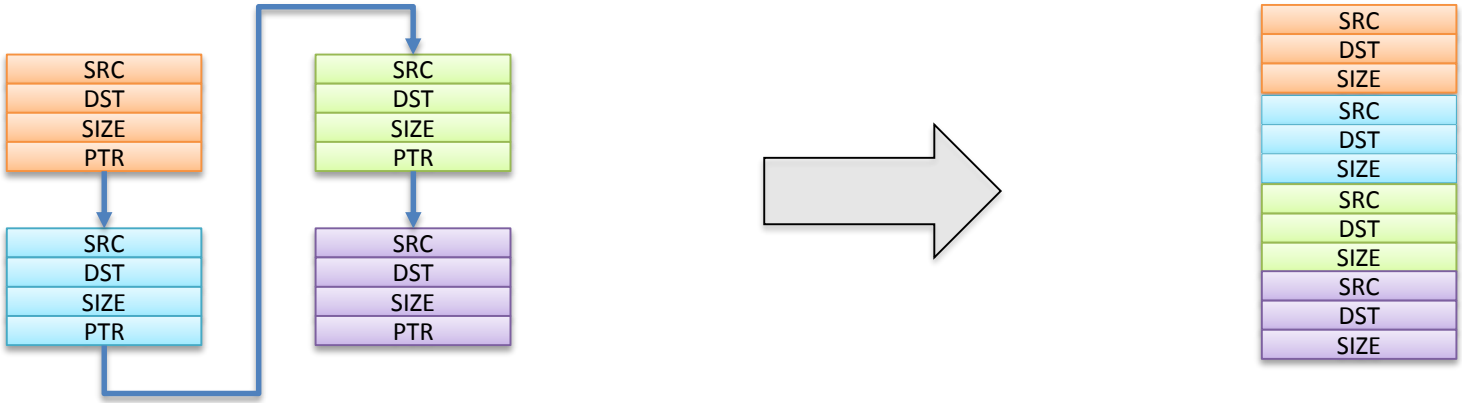


Fig. 2 - 2 outstanding requests, small packet

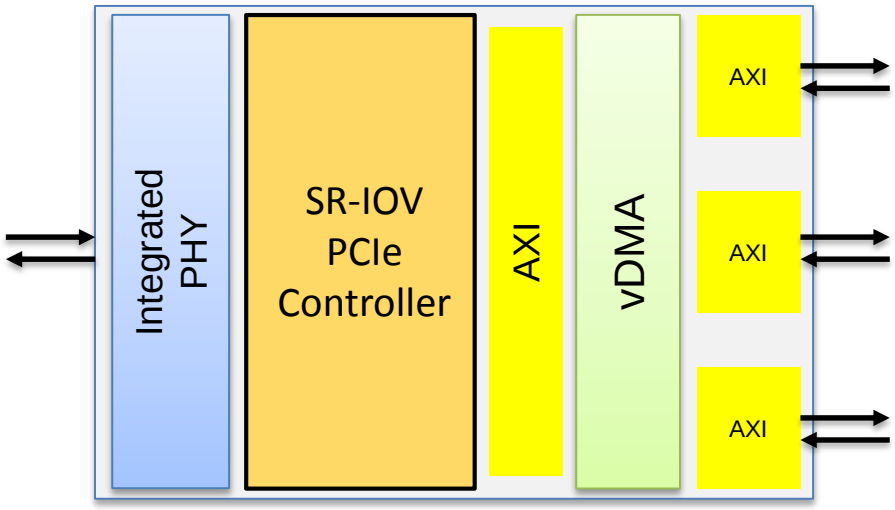
- **Increase N**



- **Propose a contiguous DMA descriptor API**

Copyright © 2017 PCI-SIG® - All Rights Reserved

The vDMA IP



- **Integrated SR-IOV PCIe controller+PHY**
 - FPGA: FPGA built-in PHY
 - ASIC: Partners PHY
- **PCIe 3.1 & 4.0 rev. 0.7**
- **2048 DMA channels**
- **512 VF**
- **256-bit AXI data path**
 - PCIe 3.1 x8
 - PCIe 4.0 x4

Copyright © 2017 PCI-SIG® - All Rights Reserved

In the Works…

- **QoS (Traffic Policies)**
 - Enhanced VM priority (ex. Weighted Round Robin)
- **512-bit architecture**
 - Support for 16GT/s x8, 8GT/s x16 throughput
- **Data integrity**
 - ECC for data path error correction
- **Streaming**
 - AMBA AXI streams

Copyright © 2017 PCI-SIG® - All Rights Reserved

Thank you for attending the PCI-SIG Developers Conference Asia-Pacific Tour 2017.

For more information please go to www.pcisig.com

Copyright © 2017 PCI-SIG® - All Rights Reserved